

# Summary of ARC's SEDAC DAAC Metadata Findings

*April 21st, 2017*

This report outlines the ARC team's metadata findings for the SEDAC DAAC. The findings reported below are present in the majority of the collection level metadata records. Detailed reports for individual collection and granule level metadata records are provided separately. The detailed reports will be used for tracking the DAAC's progress towards working off the reported findings, and metrics will be generated from the detailed reports. An explanation of the color coding found in the detailed reports is included in this report along with preliminary metrics.

## **1. Collection Level Findings**

- i. In the ECHO 10 collection level metadata records, the description includes several line breaks that interrupt the flow of the text. These line breaks are also visible in the Earthdata Search client interface. Please remove these line breaks to improve the readability of the abstract.
- ii. The 'Collection Progress' element is now a required element in UMM and is meant to describe the production status of the collection. There are three responses for the 'Collection Progress' element and those responses are chosen from a controlled vocabulary list (The vocabulary list can be found in the [UMM-C documentation](#)). In order to comply with the controlled vocabulary list, ARC recommends that the "Final" response in the ECHO 10 collection level records be updated to "COMPLETE" in SEDAC's collection level metadata.
- iii. Data format information is critical in determining data usability. ARC recommends including data format information whenever possible. When multiple data formats are provided, the data format information should be provided as separate elements in the metadata, instead of comma separated values in the same element, to support future Earthdata search functionalities.
- iv. CMR uses natural language processing (NLP) to extract some information about spatial names. Therefore some spatial keywords, such as country and state names, are not as essential for search and discovery. Please consider listing "Global" or "Global Land" as a spatial keyword rather than a list of individual regions and/or countries.  
However, if spatial keywords are an important piece of SEDAC's metadata locally, please replace invalid keywords with an appropriate keyword from

the GCMD locations list since not all spatial keywords are compliant with GCMD vocabulary.

- v. In the ECHO 10 collection level records, temporal 'Precision of Seconds' information is provided. This element is meant to provide the precision of seconds used in the measurement of the data and may not be relevant to many of SEDAC's datasets. ARC recommends removing this element from the metadata since it is not required.
- vi. In the ECHO 10 collection level records, temporal 'Ends at Present Flag' information is provided but most values are flagged as 'true.' This flag appears to be inaccurate since most data sets are no longer ongoing. ARC recommends either changing the flag to "false" or removing this value from the metadata since it is not a required element.
- vii. The Data Center Roles information (Contacts/ Contact/ Role element) is now selected from a controlled vocabulary list (The vocabulary list can be found on page 22 in the [UMM-C documentation](#)). The ARC team recommends that 'Archive' be updated to 'Archiver' in SEDAC's ECHO 10 collection level records to match the controlled vocabulary.
- viii. For the ECHO 10 records, ARC recommends that the SEDAC User Services information be placed in the organization name field and that "SEDAC User Services" be removed from the contact person name fields. Currently the 'Organization Name' field reads "Socioeconomic Data and Applications Center (SEDAC)." This can be updated to "Socioeconomic Data and Applications Center (SEDAC) User Services" to be more specific if desired. Similarly, for the DIF 10 records, ARC recommends that the SEDAC USER SERVICES contact information be used to populate the 'Contact Group' fields rather than the 'Contact Person' fields.
- ix. The platform short name (Platform/ ShortName) is a required field. For SEDAC's ECHO 10 records, ARC recommends that at least one platform short name be listed in this field. The listed platform short name should comply with the [Platforms/Sources GCMD](#) vocabulary list. Note that "Not Applicable" or "Model" is an available option.
- x. The instrument short name (Platform/ Instrument/ ShortName) is a required field. For SEDAC's ECHO 10 records, ARC recommends that at least one instrument short name be listed in this field. The listed instrument short name should comply with the [Instruments/Sensors GCMD](#) vocabulary list. Note that "Not Applicable" is an available option.
- i. Whenever feasible, there should be an Online Access URL in the metadata which points as directly as possible to the described data. This link should run through URS. Currently there are multiple Online Access URLs which

may point to the data download page on SEDAC's website, the web map service (WMS) for the collection, and/or a link to download associated maps. ARC recommends moving all of these links to the 'Online Resource/ URL' element in the ECHO10 metadata format. The only Online Access URL should be the direct download link.

- ii. Online Resource URLs must be given a URL Type. This type is controlled by the URL Content Type [GCMD vocabulary](#). Please see the detailed reports for ARC recommendations concerning URL types for each link.
- iii. The BEDI stoplight charts indicate that GIBS imagery will be created for many of SEDAC's BEDI records. ARC recommends adding links to the GIBS imagery to the collection level metadata records when the layers become available. Also, ARC recommends that links to any other services that are developed for the BEDI datasets (OPeNDAP, WMS, etc...) be included in the collection level metadata records.
- iv. Mime type information is recommended for all links to Web Map Services (WMS) and any other service links provided in the metadata. This should be done once the URL has been moved from an online access URL to an online resource URL.
- v. Many of SEDAC's datasets have spatial reference information in the metadata section of the dataset's web page. Whenever possible, ARC strongly recommends including the information about the datum and geographic coordinate system associated with the data in the metadata that is provided to CMR.
- vi. For the DIF 10 records, 'product\_level\_ID' or processing level information is missing. Please note that this is a required element. Data processing level descriptions can be found here:  
<https://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/>

## 2. Granule Level Findings

Not applicable

## 3. Explanation of Color Coding Provided in Detailed Reports

Color	Definition
Cyan	Required field based on UMM-C
Light Purple	An optional primary element with

	required sub-elements based on UMM-C
Purple	A sub-element which is only required if any information is provided in the scope of the primary element based on UMM-C
White	Completely optional field
Red	Correcting these issues should be of the highest priority
Yellow	Correcting these errors are strongly recommended but are not required
Blue	Minor error/ inconsistency; points out features noticed by the ARC Team which may help improve the robustness of the metadata but are not required to be addressed
*	Any field with an asterisk is controlled by GCMD vocabulary

#### 4. Metrics

49 Collection level records checked

<i>Collection Level</i>	# Red fields	# Yellow fields	# Blue fields	Total # fields checked
	496	640	346	3,691
	13.4%	17.3%	9.4%	